

关于稠密语言的析取结构*

郭聿琦 李 廉 许光午

(兰州大学数学力学系)

摘 要

本文利用析取语言来刻画稠密语言,得到如下结论:

1. 关于字母表 $\Sigma = \{a\}$, 其上任一稠密语言可表为 Σ 上无限多个析取语言的无交并.
2. 关于任意字母表 Σ , 其上任一稠密语言可表为 Σ 上一离散析取语言与一析取语言的无交并.

一、引言与结果陈述

令 Σ 为一有限集合, Σ^* 为由 Σ 生成的自由么半群 (free monoid). Σ 的元素称为字母, Σ 称为字母表. Σ^* 的元素和子集分别称为 Σ 上的字和语言. 关于 Σ 上任一字 x , 用 $|x|$ 表示 x 含 Σ 中字母的个数, 称为 x 的长度.

若 L 为 Σ 上一语言, 则称其为 Σ 上的①稠密语言, ②离散语言和③析取语言, 如果它依次分别满足下述条件:

- ① 关于 $w \in \Sigma^*$, 常有 $L \cap \Sigma^* w \Sigma^* \neq \emptyset$;
- ② $x, y \in L$, 且 $|x| = |y|$ 意味着 $x = y$;
- ③ Σ^* 上的同余 P_L 为 Σ^* 上的恒等, 其中 P_L 为:

$$x \equiv y (P_L), \text{ 当且仅当关于 } u, z \in \Sigma^*, \text{ 常有 } xz \in L \iff yz \in L.$$

关于上述三概念的关系的讨论, 就 $|\Sigma| \geq 2$ 的情形, 已有如下事实:

- (i) 若 L 为 Σ 上的离散语言, 则 L 在 Σ 上稠密, 当且仅当 L 在 Σ 上析取^[1].
- (ii) L 为 Σ 上的稠密语言, 当且仅当 L 包含 Σ 上一析取语言^[1].
- (iii) Σ^* 的单侧理想 I ($I = \Sigma^* I$ 或 $I \Sigma^*$. 这是一类稠密语言) 可表为 Σ 上无限个析取语言的无交并^[2,3].

更一般地, 我们与 Shyr 同时得到了

- (iv) Σ 上任一稠密语言可表为 Σ 无上限个析取语言的无交并^[4,5].

另外, 我们还曾得到

- (v) Σ 上任一稠密语言可表为 Σ 上二 (从而任意有限个) 析取语言的无交并^[5].

由 (v) 的证明^[5] 可知, 任一稠密语言事实上可表为一半离散析取语言与一析取语言的无

本文 1983 年 10 月 25 日收到.

* 中国科学院科学基金资助的课题.

交并. 所谓 Σ 上的语言 L 是半离散的, 指的是存在自然数 k , 使得 L 所含任何长度的字都不超过 k 个^[6]. 显然, 半离散性是离散性的推广^[6].

本文的主要结果是下述二定理.

定理 1. 事实 (iv) (从而事实 (iii)) 关于 $|\Sigma| = 1$ 也是对的.

于是, 事实 (iv) (包括 (iii)) 由 $|\Sigma| \geq 2$ 的情形推广到了任意 Σ .

定理 2. 关于任意字母表 Σ , Σ 上任一稠密语言可表为 Σ 上一离散析取语言与一析取语言之无交并.

于是, 一方面, 事实 (v) 由 $|\Sigma| \geq 2$ 的情形推广到了任意 Σ ; 另一方面, 事实 (v) 的结论由一析取并因子的 k -离散改进为离散.

由定理 1, 2, 事实 (ii) 随之也由 $|\Sigma| \geq 2$ 的情形推广到了任意 Σ .

但事实 (i) 不能得到事实 (ii)——(v) 那样的推广, 因为当 $|\Sigma| = 1$ 时, 语言的稠密性与无限性是一致的, 而这一情形的无限语言却分为析取与正则两类.

二、定理 1 的证明

引理 1^[7]. 设 $\Sigma = \{a\}$, L 为 Σ 上的语言. 若关于任意整数 $m > 0$, 存在 $a^{t_1}, a^{t_2} \in L$, 使得 $t_2 - t_1 > m$, 且 L 中不再含有长度介于 t_1 与 t_2 之间的字, 则 L 不是正则的 (从而是析取的).

下面给出定理 1 的证明. 令 D 为 $\Sigma = \{a\}$ 上的稠密语言. 由 D 稠密, 知 $|D| = \infty$, 因此可取到 D 的一个无限子集 $\{y_1, y_2, y_3, \dots\}$, y_1 为 D 中的最短字, 且 $|y_i| - |y_{i-1}| > i$, $i = 2, 3, \dots$. 关于每个 y_i , 作

$$T_i = \{y \in D \mid |y_{i-1}| \leq |y| < |y_i|\}, \quad i = 2, 3, \dots$$

令 N 为自然数的集合, 且

$$M_j = \{n \in N \mid n \text{ 恰有 } j \text{ 个素因子}\}, \quad j = 1, 2, \dots$$

对每个 M_j , 作

$$D_j = \bigcup_{k \in M_j} T_k.$$

显然关于任何 $k_1, k_2 \in N$, $k_1 \neq k_2$ 时, $T_{k_1} \cap T_{k_2} = \emptyset$, 因此当 $j_1 \neq j_2$ 时, $D_{j_1} \cap D_{j_2} = \emptyset$. 由上面的作法可知, 关于每一 $x \in D$, 必有 $k \in N$, 使得 $x \in T_k$, 所以

$$D = \bigcup_{j=1}^{\infty} D_j.$$

下证每个 D_j 都是 $\Sigma = \{a\}$ 上的析取语言, $j = 1, 2, \dots$. 关于任何整数 $m > 0$, 当然可取到整数 $n > m$, 且

1°. $n \in M_j$,

2°. 存在 $n_0 \in M_j$, 使得 $n_0 < n$.

于是, $T_n \cap D_j = \emptyset$. 令 $r_1 = \min\{r \in M_j \mid r > n\}$, $r_2 = \max\{r \in M_j \mid r < n\}$, 显然, $T_{r_1} \subseteq D_j$, $T_{r_2} \subseteq D_j$. 设 a^{s_1} 为 T_{r_1} 中的最短字, a^{s_2} 为 T_{r_2} 中的最长字, 由 D_j 的作法, 知 D_j 中不再含有长度介于 s_1 与 s_2 之间的字. 但由 $a^{s_1} \in T_{r_1}$, 知 $s_1 \geq |y_n|$, 由 $a^{s_2} \in T_{r_2}$, 知 $s_2 < |y_{n-1}|$, 因此, $s_1 - s_2 > |y_n| - |y_{n-1}| > n > m$. 据引理 1, D_j 是析取的. ||

三、定理 2 的证明

引理 2^[1]. 设 $|\Sigma| \geq 2$, L 为 Σ 上的语言. 则 L 是析取的, 当且仅当关于 $|u| = |v|$ 的 u, v , 常有, $u \equiv v(P_L)$ 意味着 $u = v$.

易知, 还有如下的

(*) **事实**. 令 $S_1, S_2, \dots, S_m$ 为集合 S 的 $m \geq 2$ 个子集, 且 $|S_m| \geq m$ (即 S_m 含元素个数不小于 m), 则有 $a \in S_m$, 使 $S_m \setminus \{a\}$ 与 $S_1, S_2, \dots, S_{m-1}$ 都不相同.

这是因为, 如果不然, 令 $\{a_1, a_2, \dots, a_m\} \subseteq S_m$, 则有 $S_m \setminus \{a_i\} \in \{S_k | k = 1, 2, \dots, m-1\}$, $i = 1, 2, \dots, m$, 但

$$S_m \setminus \{a_i\} \neq S_m \setminus \{a_j\}, \quad i, j = 1, 2, \dots, m, \quad i \neq j,$$

这是一个矛盾.

下面给出定理 2 的证明.

令 D 为一稠密语言, 我们将从 D 中取出一个离散析取子集 D_1 , 且使 $D_2 = D \setminus D_1$ 析取, 以完成定理的证明.

先考虑 $|\Sigma| \geq 2$ 的情形. 令 $\leq$ 为 Σ^* 上一严格序, 即 $\leq$ 为 Σ^* 上的一个全序, 且 $|x| < |y|$ 时, 常有 $x < y$. 例如, $\leq$ 可取为词汇序. 记

$$\Sigma^* = \{x_1, x_2, \dots, x_n, \dots\},$$

其中, $i < j$ 时 $x_i < x_j$. 又令

$$A_i = \{(u, v) \in \Sigma^* \times \Sigma^* | ux_i v \in D\}, \quad i = 1, 2, \dots.$$

由 D 稠密, 知 $|A_i| = \infty^{[1]}$, 即 A_i 含有无穷多个元素, $i = 1, 2, \dots$.

为构造 D_1 , 我们先从 A_1 中任取 (u_1, v_1) ; 令 $T_1 = \{(u, v) | |u| + |v| \leq |u_1| + |v_1|\}$, 再从 $A_2 \setminus T_1$ 中取一 (u'_2, v'_2) , 使得

$$A_2 \setminus (T_1 \cup \{(u'_2, v'_2)\}) \neq A_1 \setminus T_1. \quad (1)$$

这是可以作到的, 因为 $|A_1| = \infty$, $|A_2 \setminus T_1| = \infty$, 而且有上面的事实 (*). 由 (1) 式, 知有 (u_{21}, v_{21}) 属于且只属于 $A_2 \setminus (T_1 \cup \{(u'_2, v'_2)\})$ 与 $A_1 \setminus T_1$ 中的一个, 这样的元称为 $A_2 \setminus (T_1 \cup \{(u'_2, v'_2)\})$ 与 $A_1 \setminus T_1$ 的**示差元**. 如果 $(u_{21}, v_{21}) = (u'_2, v'_2)$, 那么我们令 $(u_2, v_2) = (u'_2, v'_2)$; 否则, 据事实 (*), 可取到 $(u_2, v_2) \in A_2 \setminus T_1$, 使得

$$|u_2| + |v_2| > \max\{|u'_2| + |v'_2|, |u_{21}| + |v_{21}|\},$$

$$A_2 \setminus (T_1 \cup \{(u_2, v_2)\}) \neq A_1 \setminus T_1.$$

且易知, (u_{21}, v_{21}) 也是 $A_2 \setminus (T_1 \cup \{(u_2, v_2)\})$ 与 $A_1 \setminus T_1$ 的示差元.

今若关于某一 $k \geq 2$, $A_k \setminus T_{k-1}$ 中的 (u_k, v_k) 已经取出, 使得

$$A_k \setminus (T_{k-1} \cup \{(u_k, v_k)\}) \neq A_i \setminus T_{k-1}, \quad i = 1, 2, \dots, k-1,$$

其中 $T_{k-1} = \{(u, v) | |u| + |v| \leq |u_{k-1}| + |v_{k-1}|\}$. 同时也已取出 $A_k \setminus (T_{k-1} \cup \{(u_k, v_k)\})$ 与 $A_i \setminus T_{k-1}$ 的示差元 (u_{ki}, v_{ki}) , $i = 1, 2, \dots, k-1$, 使得关于任一 $i \leq k-1$, 或者

$$(u_{ki}, v_{ki}) = (u_k, v_k),$$

或者

$$|u_{ki}| + |v_{ki}| < |u_k| + |v_k|.$$

令 $T_k = \{(u, v) \mid |u| + |v| \leq |u_k| + |v_k|\}$. 由 $|A_{k+1} \setminus T_k| = \infty$, 据事实(*), 可取到 $(u'_{k+1}, v'_{k+1}) \in A_{k+1} \setminus T_k$, 使得

$$A_{k+1} \setminus (T_k \cup \{(u'_{k+1}, v'_{k+1})\}) \neq A_i \setminus T_k, \quad i = 1, 2, \dots, k.$$

如果关于 $i_0 \leq k$, $A_{k+1} \setminus T_k = A_{i_0} \setminus T_k$, 那么 $A_{k+1} \setminus T_k$ 中任何元 (u, v) 都是 $A_{k+1} \setminus (T_k \cup \{(u, v)\})$ 与 $A_{i_0} \setminus T_k$ 的示差元. 关于使 $A_{k+1} \setminus T_k \neq A_{i_t} \setminus T_k$ 的 i_t , $i_t \leq k$, $t = 1, 2, \dots, h$, 可取到 $(u_{k+1, i_t}, v_{k+1, i_t})$ 作为 $A_{k+1} \setminus (T_k \cup \{(u'_{k+1}, v'_{k+1})\})$ 与 $A_{i_t} \setminus T_k$ 的示差元, 使得 $(u_{k+1, i_t}, v_{k+1, i_t}) \neq (u'_{k+1}, v'_{k+1})$. 据事实(*), 可取到 $(u_{k+1}, v_{k+1}) \in A_{k+1} \setminus T_k$, 使得

$$|u_{k+1}| + |v_{k+1}| > \max\{|u_{k+1, i_t}| + |v_{k+1, i_t}| \mid t = 1, 2, \dots, h\} \cup \{|u'_{k+1}| + |v'_{k+1}|\},$$

$$A_{k+1} \setminus (T_k \cup \{(u_{k+1}, v_{k+1})\}) \neq A_i \setminus T_k, \quad i = 1, 2, \dots, k.$$

易知, $(u_{k+1, i_t}, v_{k+1, i_t})$ 也是 $A_{k+1} \setminus (T_k \cup \{(u_{k+1}, v_{k+1})\})$ 与 $A_{i_t} \setminus T_k$ 的示差元, $t = 1, 2, \dots, h$. 而关于 $i \leq k$, $i \neq i_t$, $t = 1, \dots, h$, 令 $(u_{k+1, i}, v_{k+1, i}) = (u_{k+1}, v_{k+1})$, 它为 $A_{k+1} \setminus (T_k \cup \{(u_{k+1}, v_{k+1})\})$ 与 $A_i \setminus T_k$ 的唯一示差元.

于是, 由归纳法得到 $\Sigma^* \times \Sigma^*$ 的无限子集

$$T = \{(u_i, v_i) \mid i = 1, 2, \dots\}.$$

作 Σ 上语言

$$D_1 = \{u_i x_i v_i \mid i = 1, 2, \dots\},$$

由 A_i 的意义及 $(u_i, v_i) \in A_i$, 显然 D_1 为 D 的子集. 由 (u_i, v_i) 的取法, 有

$$|u_{i_2}| + |v_{i_2}| > |u_{i_1}| + |v_{i_1}|, \quad i_2 > i_1.$$

由 $\leq$ 的严格性, 有

$$|x_{i_2}| \geq |x_{i_1}|, \quad i_2 > i_1.$$

因此,

$$|u_{i_2} x_{i_2} v_{i_2}| > |u_{i_1} x_{i_1} v_{i_1}|, \quad i_2 > i_1.$$

从而, D_1 离散. D_1 的稠密性由其构造立得. 于是, 据事实(i), D_1 为(离散)析取语言.

下证 $D \setminus D_1$ 的析取性. 关于任意 x_i, x_j , $1 \leq i < j$, $|x_i| = |x_j|$, 由 (u_{ji}, v_{ji}) 为

$$A_j \setminus (T_{j-1} \cup \{(u_j, v_j)\}) = A'_j$$

与

$$A_i \setminus T_{j-1} = A'_i$$

的示差元, 知 (u_{ji}, v_{ji}) 属于且仅属于 A'_j 与 A'_i 中的一个.

(a) 当 $(u_{ji}, v_{ji}) \in A'_j$, $(u_{ji}, v_{ji}) \notin A'_i$ 时, 有

$$u_{ji} x_j v_{ji} \in D \setminus D_1^{(j-1)},$$

$$u_{ji} x_i v_{ji} \notin D \setminus D_1^{(j-1)},$$

其中

$$D_1^{(j-1)} = \{w \in D_1 \mid |w| \leq |u_{j-1} x_{j-1} v_{j-1}|\}.$$

从而有

$$u_{ji} x_j v_{ji} \in D \setminus D_1, \quad (2)$$

$$u_{ji} x_i v_{ji} \notin D \setminus D_1. \quad (3)$$

事实上, 由 $D_1 \supseteq D_1^{(j-1)}$ 显然(3)式成立; (2)式成立是因为, $r < j$ 时, $|u_{ji}| + |v_{ji}| > |u_r| + |v_r|$, $r > j$ 时, $|u_r| + |v_r| > |u_{ji}| + |v_{ji}|$, 因此, 在这些情形, 必有 $u_{ji} x_j v_{ji} \neq u_r x_r v_r$;

而 $r = j$ 时, 若 $u_{ji}x_jv_{ji} = u_jx_jv_j$, 则有 $|u_{ji}| + |v_{ji}| = |u_j| + |v_j|$, 但 $(u_{ji}, v_{ji}) \in A'_j$, 这与 $|u_j| + |v_j| > |u_{ji}| + |v_{ji}|$ 矛盾.

(b) 当 $(u_{ji}, v_{ji}) \notin A'_j, (u_{ji}, v_{ji}) \in A'_i$ 时, 有

$$u_{ji}x_jv_{ji} \notin D \setminus D_1, \quad (2')$$

$$u_{ji}x_jv_{ji} \in D \setminus D_1. \quad (3')$$

事实上, $(u_{ji}, v_{ji}) \notin A_j$ 时, 当然 (2') 式成立; 而 $(u_{ji}, v_{ji}) \in A_j$ 时, 由 $(u_{ji}, v_{ji}) \notin A'_j$, 必有 $(u_{ji}, v_{ji}) \in (T_{j-1} \cup \{(u_j, v_j)\})$, 但由 $(u_{ji}, v_{ji}) \in A'_i$, 知 $(u_{ji}, v_{ji}) \notin T_{j-1}$, 因此, $(u_{ji}, v_{ji}) = (u_j, v_j)$, 从而 $u_{ji}x_jv_{ji} = u_jx_jv_j \in D_1$, 即也有 (2') 式成立. 要证 (3') 式成立, 由 $(u_{ji}, v_{ji}) \in A'_i$, 知 $u_{ji}x_iv_{ji} \in D$, 只须再证得 $u_{ji}x_iv_{ji} \notin D_1$ 好了. 若 $u_{ji}x_iv_{ji} \in D_1$, 则有 $u_{ji}x_iv_{ji} = u_r x_r v_r$, 同 (2) 式的讨论, 必有 $r = j$, 即 $u_{ji}x_iv_{ji} = u_jx_jv_j$, 但 $|x_j| = |x_i|$, 因此有 $|u_{ji}| + |v_{ji}| = |u_j| + |v_j|$, 又由 (u_j, v_j) 的取法, 必有 $(u_{ji}, v_{ji}) = (u_j, v_j)$, 即 $u_jx_jv_j = u_jx_iv_j$, 从而 $x_i = x_j$, 这与 $x_i \neq x_j$ 矛盾, 于是 $u_{ji}x_iv_{ji} \notin D_1$.

(2), (3) 式和 (2'), (3') 式证明

$$x_i \not\equiv x_j \quad (P_{D \setminus D_1}),$$

由 i, j 的任意性, 据引理 2 知, $P_{D \setminus D_1}$ 为恒等关系. 故 $D \setminus D_1$ 为析取语言.

再考虑 $|\Sigma| = 1$ 的情形. 设 $\Sigma = \{a\}$. 由 D 稠密, 知 D 为无限语言, 令 $D = \{a^{i_1}, a^{i_2}, \dots\}$, $i_1 < i_2 < \dots$.

作

$$P_k = \{a^{i_2^k}, a^{i_2^k+1}, \dots, a^{i_2^{k+1}} - 1\}, \quad k = 0, 1, 2, \dots,$$

显然 D 为这些 P_k 的无交并.

$$\text{令 } D_1 = \bigcup_{l=0}^{\infty} P_{2l}, \text{ 则 } D_2 = D \setminus D_1 = \bigcup_{l=0}^{\infty} P_{2l+1}.$$

下证 D_1, D_2 析取.

关于任一自然数 m , 存在自然数 t , 使得 $2^t > m$. 因为 $a^{i_2^{2t+1}-1}, a^{i_2^{2t+2}} \in D_1$, 并且 D_1 中没有长度介于 $i_2^{2t+1}-1$ 与 i_2^{2t+2} 之间的字, 又因为 $i_2^{2t+2} - i_2^{2t+1}-1 \geq 2^{2t+2} - 2^{2t+1} + 1 = 2^{2t+1} + 1 > 2^t > m$, 所以, 据引理 1, 知 D_1 是析取的. 同样可证, D_2 析取. 此时, 显然, D_1 与 D_2 都是离散的.

参 考 文 献

- [1] Reis, C. M. and Shyr, H. J., Some Properties of Disjunctive Languages on a Free Monoid, *Inform. Contr.*, 37 (1978), 334—344.
- [2] Shyr, H. J., Disjunctive Languages on a Free Monoid, *Inform. Contr.*, 34(1977), 123—129.
- [3] Shyr, H. J. and Thierrin, G., Disjunctive Languages and Codes, *Fundamentals of Computation Theory, Proceedings of the 1977 Inter. ECT-Conference, Poznan, Poland, Lecture Notes in Computer Science* N. 56, Springer-Verlag, 1977, 171—176.
- [4] Thierrin, G., *Private Communication*, 1983.
- [5] 许光午、郭聿琦、李廉, 稠密语言关于析取语言的并分解(待发表)
- [6] Kunze, M., Shyr, H. J. and Thierrin, G., H-Bounded and Semi-discrete Languages, *Inform. Contr.*, 51 (1981), 2: 174—187.
- [7] Eilenberg, S., *Automata, Languages, and Machines*, Academic Press, New York and London, 1974.